IEEE Transactions on Communications, Volume 68, Issue 3, March 2020, pp. 1479 - 1493
DOI:10.1109/TCOMM.2019.2959354

High-dimensional Metric Combining for Non-coherent Molecular Signal Detection

Zhuangkun Wei, Weisi Guo, Bin Li, Jerome Charmet, and Chenglin Zhao

***Abstract*—In emerging Internet-of-Nano-Thing (IoNT), information will be embedded and conveyed in the form of molecules through complex and diffusive medias. One main challenge lies in the long-tail nature of the channel response causing inter-symbol-interference (ISI), which deteriorates the detection performance. If the channel is unknown, existing coherent schemes (e.g., the state-of-the-art maximum a *posteriori*, MAP) have to pursue complex channel estimation and ISI mitigation techniques, which will result in either high computational complexity, or poor estimation accuracy that will hinder the detection performance. In this paper, we develop a novel high-dimensional non-coherent detection scheme for molecular signals. We achieve this in a higher-dimensional metric space by combining different non-coherent metrics that exploit the transient features of the signals. By deducing the theoretical bit error rate (BER) for any constructed high-dimensional non-coherent metric, we prove that, higher dimensionality always achieves a lower BER in the same sample space, at the expense of higher complexity on computing the multivariate posterior densities. The realization of this high-dimensional non-coherent scheme is resorting to the Parzen window technique based probabilistic neural network (Parzen-PNN), given its ability to approximate the multivariate posterior densities by taking the previous detection results into a channel-independent Gaussian Parzen window, thereby avoiding the complex channel estimations. The complexity of the posterior computation is shared by the parallel implementation of the Parzen-PNN. Numerical simulations demonstrate that our proposed scheme can gain 10dB in SNR given a fixed BER as 10^{-4}, in comparison with other state-of-the-art methods.**

***Index Terms*—Non-coherent detection, high-dimensional metric, Bayessian rule, Parzen-probabilistic neural network, machine learning.**

I. Introduction

RECENT advances in nano-technology have attracted widespread interests in a range of applications, including the Internet-of-Nano-Things (IoNT) [1], [2], precision medicine [3], and covert signalling in electromagnetically denied environments. To support these, increasing efforts have been spent on the studies of molecular communications via diffusion (MCvD), whereby digital information is modulated via the chemical structures of molecules, and then undergo a combination of diffusion and advection propagation. Compared with the traditional electromagnetic waves (EMW) or acoustic waves [4], the stochastic nature of the propagation allows nano-scale information carriers to diffuse through complex medias, thereby making MCvD a promising communication candidate for bioengineering applications.

In the context of the signal detection in the MCvD, three challenges should be considered. First, the long-tail nature of the channel response causes severe inter-symbol-interference (ISI), which will deteriorate the accuracy of signal detection. Second, the underlying diffusion model will be inaccurate (or even unavailable) when dealing with complex channels (e.g., a micro-fluidic and absorbing channels with complex and irregular shapes obstacles [5]–[9]). In this view, understanding a total channel state information (CSI) requires a large amounts of resources, therefore making those model-related detection schemes less practical. Third, a nano-receiver aiming at measuring the number of target molecules should be subject to a limited energy expenditure and also ensure real-time communication.

A. Related Works

The pure mass diffusion channel impulse response (CIR) and the noise distribution formulation are well established (even for certain boundary and reaction conditions). As such, there are significant efforts on developing coherent signal processing schemes. Such coherent schemes pursue signal detection, leveraged on the estimated CSI to compensate the channel effect (e.g., the ISI). For instance, we can list various coherent detection as well as channel estimation algorithms in [10]–[13]. The state-of-the-art maximum a *posteriori* (MAP) [11] uses a designed pilot sequence provided from [10] to estimate the CSI and then relies on the MAP concept to detect molecular signals. However, these coherent inference schemes have two drawbacks. First, the coherent schemes relying on the CSI acquisitions in [10] are only practical when the coherence time of the MCvD channels are relatively large [14]. Otherwise, for the scenarios where the channels are time-varying rapidly, they have to repeatedly track the CSI variations in order to ensure the detection accuracy, which inevitably causes long pilot sequence overhead, and substantial energy consumption. Second, the coherent schemes are sensitive to the erroneous estimations of the CSI. Such errors may lead to the unreliable computations of the posterior densities, which if used for the coherent detection process (e.g., MAP), may result in low detection accuracy (as is shown in the Figs. 7-8 in the Simulation section).

Alternatively, the other group of detection method is referred to as non-coherent schemes, which do not rely on the availability of the CSI for signal detection. Such schemes can be assisted by the ISI mitigated methods such as enzyme equalizer [15], [16], and the stochastic resonance [17].

Zhuangkun Wei, Weisi Guo and Jerome Charmet are with the University of Warwick, Coventry, West Midlands, CV4 7AL, UK. (Email: Zhuangkun.wei@warwick.ac.uk)

Bin Li and Chenglin Zhao are with the School of Information and Communication Engineering (SICE), Beijing University of Posts and Telecommunications (BUPT), Beijing, 100876, China.

The popular non-coherent schemes for MCvD are listed in [14], [17]–[20]. We analyze the advantages and drawbacks of these papers in the following. The optimal non-coherent maximum likelihood (ML) detectors (for multiple-symbol MS, and symbol-by-symbol SS) have been proposed and studied in [14], which considered an accurate Poisson model, and provided optimal (or suboptimal) detection algorithms for MCvD. The results showed a promisingly lower bit error rate (BER) in the case of the negligible ISI (which can be obtained by the enzyme equalizer in [15], [16]). The minor challenge lies in that, for the case where ISI is intensive, the detection accuracy saturates to a BER floor, suggesting the limitation to strong ISI scenarios. Then, in [18], the constant-composition codes have been proposed to further address the ISI issue of the non-coherent ML detectors, at the expense of a decrease in data rate, as it introduces extra bits for coding. In our previous papers [17], [19], [20], we proposed three metrics for non-coherent detectors, aiming to address the ISI issue by exploiting the transient features of molecular signals that are insensitive to ISI. Premised on the metrics, these non-coherent detectors transform the detection issue into an two-hypothesis test, thereby providing a new paradigm for signal detection in MCvD with non-trivial ISI. However, such metric based non-coherent detectors linearly combine the three metrics, which does not fully employ the transient features and results in a loss of signal-to-noise ratio (SNR). Also, there is still a lack of understanding on the theoretical performance bound of the proposed metric based non-coherent detectors. These two open challenges constitute the motivation to extend our previous researches.

B. Contributions

In this work, we suggest a novel non-coherent detection scheme for MCvD, based on a newly designed high-dimensional metric. To sum up, the main contributions of this paper are listed as follows:

(1) A high-dimensional metric is designed via an exploration of the transient features of the molecular signals. Compared with the coherent detection schemes that require the CSI to compensate the ISI effect, such feature-based metric is inherently insensitive to the ISI, therefore enabling the design of the non-coherent detection. Also, as opposed to our previously proposed linear non-coherent works in [17], [19], [20], the high-dimensional metric is capable of improving the detection accuracy, as it constructs the multivariate posterior functions, and thereby has a greater signal-to-noise ratio (SNR). The proof is in Eqs. (24)-(28).

(2) From the theoretical perspective, we compute the theoretical BER for any designed high-dimensional non-coherent scheme premised on the Bayesian inference. Specially, we prove that, as the high-dimensional non-coherent metrics are constructed by the same samples, the metric with higher dimensions has a lower BER. This further indicates that our proposed high-dimensional non-coherent scheme outperforms the previously linear ones, and that the lower-bound of BER converges to that of the coherent MAP.

(3) Given the unknown CIR, in order to approximate the posterior probability density functions (PDFs) of the high-dimensional metric, but still avoid the complex parameter estimations, we suggest the Parzen window technique based probabilistic neural network (Parzen-PNN). We prove that by adopting a Gaussian Parzen window, the Parzen-PNN can approach the theoretical BER. Other than the conventional channel estimation algorithms that suffer from the substantially computational complexity, and the long pilot sequence overhead, this Parzen-PNN directly takes the previous detection results into a channel-independent Gaussian Parzen window to construct the approximated PDFs, therefore avoiding the computational burden on CSI acquisition. It is noteworthy that when comparing with the previously proposed linear non-coherent schemes, the Parzen-PNN has no low-complexity advantages as it computes the multivariate PDFs. However, attributed to the parallel property of the Parzen-PNN, the multivariate PDF computations can be assigned into each parallel threats [21]–[23]. Therefore, it is capable of ensuring the real-time communication and the limited energy-expenditure for MCvD scenarios.

(4) We evaluate the detection performance of our proposed high-dimensional non-coherent scheme via simulations. The results demonstrate that the accuracy of signal detection is improved about 10dB in SNR given a fixed BER as BER=10^{-4}, as opposed to our previous algorithm, and the MAP with the unknown CSI. Also, the computational complexity and the storage resources are much lower than that of the coherent MAP (as is shown in Table. 1). Hence, by casting the accurate detection into an energy-efficient framework, the proposed scheme shows the great promise to the MCvD applications.

The rest of this article is structured as follows. In Section II, the system structure of the MCvD is specified, along with a short description of the state-of-the-art MAP and our previously proposed linear non-coherent schemes. In Section III, we elaborate our designed high-dimensional non-coherent scheme, compute its theoretical BER, and prove that it has a better communication accuracy than the low-dimensional non-coherent schemes. In Section IV, we introduce the Parzen-PNN with the aim of the realization of our scheme. Numerical simulations are provided in Section V. Finally, we conclude this study in Section VI along with discussions of application areas and future impact.

II. System Models

A. Molecular Communications

Similar to an EMW-based communication system, a generalized model for MCvD is illustrated in Fig. 1, consisting of a nano-transmitter, a propagation channel, and a nano-receiver.

1) Nano-transmitter: The nano-transmitter may be either a single cell/organism in a biological system, or an artificially designed hardware. The ON/OFF key (OOK) information bit $\alpha_k = j \in \{0, 1\}$ $(k = 1, \cdots, +\infty)$ can be modulated via either a number of a specific type of molecules, or the phase (i.e., interval) [24]. Here, we consider the amplitude shift key (ASK) from [13]. The modulated signal, denoted as $s(t)$ can be thereby expressed as:

$$s(t) = Q \cdot \sum_{k=0}^{\infty} \alpha_k \cdot \delta(t - kT_b), \tag{1}$$

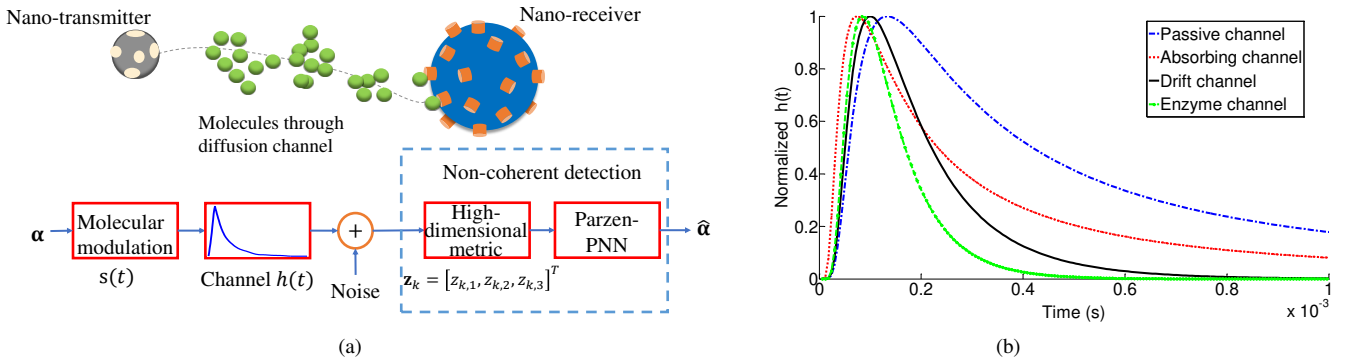

Fig. 1. Illustration of MCvD system. (a) gives the schematic flow where the binary information $\boldsymbol{\alpha} = [\alpha_0, \cdots, \alpha_k]^T$ are modulated by the Nano-transmitter via molecules, which will then propagate through the diffusive channel $h(t)$, and received by the nano-receiver. After sampling the received signal, the receiver will construct the high-dimensional metric and then rely on the Parzen-PNN algorithm to detect the information $\hat{\boldsymbol{\alpha}} = [\hat{\alpha}_0, \cdots, \hat{\alpha}_k]^T$. (b) shows the different cases of diffusive channels and their corresponding CIRs

where T_b is the bit interval, and Q represents the emitted number of the molecules. Here, $\delta(\cdot)$ accounts for the Dirac function that is adopted to describe the pulse shape.

2) Propagation channel: Typically, a diffusive channel model and its corresponding CIR (denoted as $h(t)$) remains unknown due to the infeasible computations of the complex channel properties (e.g., a mixture with absorbing and obstacles, presence of advection forces, fluctuating channel parameters...etc.). Hence, we assume an unknown CIR for our high-dimensional non-coherent signal detection, by randomly selecting one of the two widely-used diffusion models in Eq. (2) for performance analysis and comparisons with the state-of-the-art MAP. The two models (including the passive and the absorbing channels) and their CIRs are given as follows [7], [9], [15], [25]:

$$h(t) = \begin{cases} \frac{V}{(4\pi Dt)^{3/2}} \exp\left(-\frac{(r-vt)^2}{4Dt} - \beta t\right) & \text{passive} \\ \frac{R}{R+r} \frac{r}{\sqrt{4\pi Dt^3}} \exp\left(-\frac{(r-vt)^2}{4Dt} - \beta t\right) & \text{absorbing} \end{cases} \tag{2}$$

where V gives the volume of a spherical nano-receiver with its radius R, r is the transmitted distance, D represents the diffusion coefficient, v is the velocity of the drift, and β is the degradation exponent of the enzyme interactions.

As we assign various parameters for Eq. (2), we can see the different CIRs in Fig. 1(b). It is noteworthy that although the exact shapes of the CIRs depends heavily on the model expressions and their model-related parameters, there are still transient features in common, e.g., an obvious rising edge for the appearance of an 1-bit, the inflexions, and the signal-difference between two adjacent bits, which may give hints for the design of a non-coherent detection paradigm in the receiver.

3) Nano-receiver: A nano-scale receiver may include two components [26]: (i) the sensory units that sense and convert signaling molecules into quantities for detection, and (ii) the signal processing unit for signal detection. Here, we describe the formulas of the signal model of received molecules. The biological implementation of the nano-receiver will be discussed in Section IV. D.

At the receiver end, the received signal is expressed by counting the number of observed molecules. Here, a multi-sample receiver is adopted. Let T_b denote the symbol interval and T_s denote the sampling interval (subject to Nyquist theorem [11], [20] with $MT_s = T_b$). Then, the received signal at the sampling time $i \cdot T_s \in ((k-1)T_b, kT_b]$, denoted as y_i, can be expressed via the Poisson distribution, i.e. [26],

$$y_i \sim \mathcal{P}\left(\sum_{l=k-L+1}^{k} Q \cdot h_{i-(l-1)M} \cdot \alpha_l + \bar{\epsilon}_{\text{int}}\right), \tag{3}$$

where L denotes the length of the ISI, $h_i = h(iT_s)$, and $\bar{\epsilon}_{\text{int}}$ represents the mean number of molecules that originates from the external interfering noise. Deduced from Eq. (3), an additive signal model is derived, i.e. [26],

$$y_i = \underbrace{Q \cdot h_{i-(k-1)M} \cdot \alpha_k}_{\text{current signal}} + \underbrace{\sum_{l=k-L+1}^{k-1} Q \cdot h_{i-(l-1)M} \cdot \alpha_l}_{\text{ISI}} + \underbrace{\varepsilon_i + \epsilon_{\text{int}}}_{\text{noise}}. \tag{4}$$

In Eq. (4), $Qh_{i-(k-1)M} \cdot \alpha_k$ accounts for the current information signal that is required to be detected. $\sum_{l=1}^{k-1} Q \cdot h_{i-(l-1)M} \cdot \alpha_l$ gives the ISI induced by previously transmitted bits due to the long tail nature of $h(t)$. The noise component i.e., $\varepsilon_i + \epsilon_{\text{int}}$ involves the signal-dependent diffusion noise, ε_i and the signal-independent interference noise, $\epsilon_{\text{int}} \sim \mathcal{P}(\bar{\epsilon}_{\text{int}})$. The distribution of the diffusion noise component ϵ_i can be computed in accordance with Eq. (3), i.e., $\varepsilon_i \sim \mathcal{P}_0(\sum_{l=1}^{k} Qh_{i-(l-1)M} \cdot \alpha_l)$[1], which makes it signal-dependent. In the following, we consider the signal-independent noise for analysis, under the assumption that the external noise ϵ_{int} is dominant. In the simulation section (Section V), we demonstrate that the proposed high-dimensional non-coherent scheme can also achieve a reliable BER, when the noise is signal-dependent.

With the help of Eqs. (3)-(4), the purpose of this paper is to detect the current informative bit α_k from the received signal $\mathbf{y}_{1:kM} = [y_1, ..., y_{kM}]^T$ in the absence of the expression of

[1]Here, the notation $X \sim \mathcal{P}_0(\lambda)$ represents the distribution of $X = Y - \lambda$ with $Y \sim \mathcal{P}(\lambda)$.

the CIR $h(t)$, Q, and the means of the signal-dependent and signal-independent noises.

B. Two State-of-the-Art Methods for Comparison

1) Coherent MAP: In general, an MAP detector aims at maximizing *a posteriori* probability of the unknown information bits conditioned on the received samples. The coherent MAP for MCvD is provided in [11]. Denote $\tilde{h}_k = \sum_{i=(k-1)M+1}^{kM} h_i$, and $\tilde{y}_k = \sum_{i=(k-1)M+1}^{kM} y_i$. The CSI is defined as $u_k = \sum_{l=k-L+1}^{k} Q \cdot \tilde{h}_{l-k+1} \cdot \alpha_l$, whose value belongs to a size of 2^L set $\mathbb{S}$. The maximal posterior density is computed as [11], [27], [28]:

$$\begin{aligned}
&\max_{\boldsymbol{\alpha}_{1:k}} p(\boldsymbol{\alpha}_{1:k}|\tilde{\mathbf{y}}_{1:k}) = \max_{\mathbf{u}_{1:k}} p(\mathbf{u}_{1:k}|\tilde{\mathbf{y}}_{1:k}) \propto \max_{\mathbf{u}_{1:k}} p(\tilde{\mathbf{y}}_{1:k}, \mathbf{u}_{1:k}) \\
&= \max_{\mathbf{u}_{1:k}} p(u_1) p(\tilde{\mathbf{y}}_{1:k}, \mathbf{u}_{2:k}|u_1) \\
&= \max_{u_1} p(u_1) p(\tilde{y}_1|u_1) \cdot \max_{\mathbf{u}_{2:k}} p(\tilde{\mathbf{y}}_{2:k}, \mathbf{u}_{2:k}|u_1) \\
&= \max_{u_1} p(u_1) p(\tilde{y}_1|u_1) \cdot \max_{u_2} p(u_2|u_1) p(\tilde{y}_2|u_2) \cdot \cdots \cdot \max_{u_k} p(u_k|u_{k-1}) p(\tilde{y}_k|u_k)
\end{aligned} \tag{5}$$

Eq. (5) is realized via the Viterbi algorithm for state tracking [27], [28]. Then, decisions are taken by means of trace back approach. At first, for any $u_k \in \mathbb{S}$, $p(\tilde{y}_k|u_k)$ is computed. Then, for each $u_l \in \mathbb{S}, l = k, \cdots, 1$, we compute $\max_{u_l} p(u_l|u_{l-1}) p(\tilde{y}_l|u_l)$ and store the corresponding transition $u_{l-1} \rightarrow u_l$. The iteration ends at $l = 1$, with $\hat{u}_1 = \text{argmax}_{u_1 \in \mathbb{S}} p(u_1) p(\tilde{\mathbf{y}}_{1:k}, \mathbf{u}_{2:k}|u_1)$. Then, $\hat{u}_l, 2 \leq l \leq k$ can be recursively identified via the stored transitions. Finally, the corresponding bits $\boldsymbol{\alpha}_{1:k}$ are derived given $u_k = \sum_{l=k-L+1}^{k} Q \cdot \tilde{h}_{l-k+1} \cdot \alpha_l$.

It is noteworthy that, however, in the above coherent MAP, the accurate estimation of the CIR will be indispensable, as one needs to compute the likelihood density $p(\tilde{y}_l|u_l)$ and the transitional probability $p(u_{l+1}|u_l)$ for the Viterbi tracing. This will cause large expenditures of computation and storage resources [20].

2) Linear non-coherent method: In essence, the previously designed non-coherent method aims at amplifying the SNR via a linear combination of three metrics (i.e., $c_{k,1}$, $c_{k,2}$ and $c_{k,3}$) that explore the transient features of molecular signals. The signal detection process is pursued by constructing c_k^{linear}, and comparing it with a heuristic threshold η_k, i.e., [17]

$$c_k^{\text{linear}} = c_{k,1} + c_{k,2} + c_{k,3} \underset{\hat{\alpha}_k=0}{\overset{\hat{\alpha}_k=1}{\gtrless}} \eta_k. \tag{6}$$

In Eq. (6), $c_{k,1}$ is referred as the local geometry shape that characterizes the maximum inflexion induced by 1-bit, $c_{k,2}$ describes the inflexion caused by the new arrival of 1-bit, while $c_{k,3}$ gives the signal-difference between successive bits. However, two drawbacks still remain. Firstly, $c_{k,1}$ performs poorly as the ISI increases, due to the fluctuating inflexions in the presence of a strong ISI. Secondly and more importantly, the effect of the SNR amplification is not promising, due to the limitation of the 1-dimensional metric (which will be analyzed in Section III. D).

III. Theory of High-dimensional Non-coherent Detection

In this section, we elaborate the theory of a novel high-dimensional non-coherent detection scheme for MCvD. We firstly construct the high-dimensional metric by re-designing and re-combining the metrics that explore the transient features of molecular signals. Then, premised on the high-dimensional metric, the theoretical BER is deduced via the computation of the decision surface. It is noteworthy that the non-coherent detection refers to the schemes that do not rely on the availability of the CSI, which indicates the unavailability of such decision surface for detection. As such, we only use it for the computation of the theoretical BER. The realization of this high-dimensional non-coherent detection scheme, provided in Section IV, does not rely on this decision surface.

A. Construction of High-dimensional Metric

In comparison with the linear metric in Eq. (6), the d-dimensional metric denoted as $\mathbf{z}_k$ is a vector composed of d sub-metrics (e.g., the rising edge, and the successive properties between adjacent symbols) that can describe the transient features of the signal. Here, we express $\mathbf{z}_k$ via a $d \times kM$ transformation matrix $\mathbf{A}_{d \times kM}$, i.e.,

$$\mathbf{z}_k = \mathbf{A}_{d \times kM} \cdot \mathbf{y}_{1:kM}, \tag{7}$$

where each row of $\mathbf{z}_k$ denotes one sub-metric.

In the context of the MCvD, $d = 3$ sub-metrics are designed by respectively exploring **1)** the *local rising-edge* in each symbol, denoted as $z_{k,1}$, **2)** the two *successive properties* between two adjacent symbols, denoted as $z_{k,2}$ and $z_{k,3}$, i.e. ,

$$\mathbf{z}_k = [z_{k,1}, z_{k,2}, z_{k,3}]^T. \tag{8}$$

1) Local rising-edge: Taking the kth interval with $M = T_b/T_s$ samples as an example, in the case of $\alpha_k = 1$, the output y_i will have a distinct rising edge. As is shown in Fig. 2(a), this rising edge can be expressed by the difference of its maximum (computed by averaging its neighbourhood $\mathcal{N}_{\max}$) from the beginning $\mathcal{N}_0$. In practice, by specifying the widths of $\mathcal{N}_{\max}$ and $\mathcal{N}_0$ to be $|\mathcal{N}_{\max}| = |\mathcal{N}_0| = M/4$, we define the metric of the local rising-edge as:

$$z_{k,1} \triangleq \frac{1}{|\mathcal{N}_{\max}|} \sum_{i \in \mathcal{N}_{\max}} y_i - \frac{1}{|\mathcal{N}_0|} \sum_{i \in \mathcal{N}_0} y_i. \tag{9}$$

It is easily noted that in the case of $\alpha_k = 1$, $z_{k,1}$ will be larger than 0, otherwise, it will be smaller that 0 when $\alpha_k = 0$. Also, compared with the previously designed geometry shape i.e., $c_{k,1}$ in Eq. (6), the rising edge (i.e., $z_{k,1}$) will not disappear with an increasing intensity of the ISI, thereby capable of characterising the transient feature of the molecular signals in different conditions of the ISI. In these views, $z_{k,1}$ can be employed as a high-performance metric to distinguish whether there are new arrivals of molecules induced by 1-bit at the nano-receiver at current time interval.

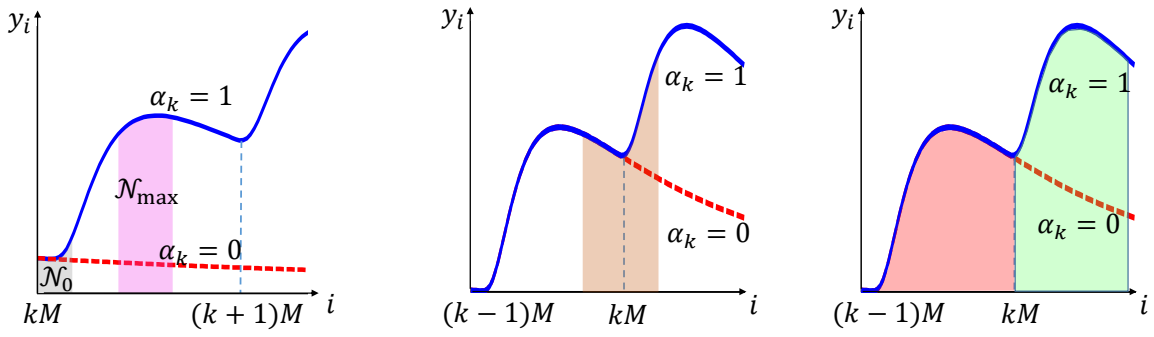

Fig. 2. Illustration of sub-metrics. (a) gives the local rising edge metric $z_{k,1}$, whilst (b) shows the successive properties that contains the inflexion metric $z_{k,2}$ and the energy-difference $z_{k,3}$

2) Successive properties: When it comes to two successive slots $k-1$ and k, the transient features will be quite different with respect to two cases (i.e., $\alpha_k = 1$ and $\alpha_k = 0$). Here, we adopt our previously designed sub-metrics (i.e., the minimum inflexion $c_{k,2}$ and the energy-difference $c_{k,3}$, as is shown in Fig. 2(b)), i.e.,

$$z_{k,2} \triangleq c_{k,2} = -\frac{1}{|\mathcal{N}_{\text{infx}}|}\cdot\sum_{i\in\mathcal{N}_{\text{infx}}} y_i + \frac{1}{2}\cdot\left(\frac{1}{|\mathcal{N}_{\text{left}}|}\sum_{i\in\mathcal{N}_{\text{left}}} y_i + \frac{1}{|\mathcal{N}_{\text{right}}|}\sum_{i\in\mathcal{N}_{\text{right}}} y_i\right), \tag{10}$$

$$z_{k,3} \triangleq c_{k,3} = \frac{1}{M}\cdot\sum_{i=kM+1}^{(k+1)M} y_i - \frac{1}{M}\cdot\sum_{i=(k-1)M+1}^{kM} y_i. \tag{11}$$

In Eqs. (10)-(11), $\mathcal{N}_{\text{infx}}$ is the set composed of the minimum inflexion and its neighbourhoods. $\mathcal{N}_{\text{left}}$ and $\mathcal{N}_{\text{right}}$ are the left and right sets of $\mathcal{N}_{\text{infx}}$. Detailed settings of $\mathcal{N}_{\text{infx}}$, $\mathcal{N}_{\text{left}}$, and $\mathcal{N}_{\text{right}}$ are provided in [17].

3) Advantages of high-dimensional metric: The high-dimensional metric in Eqs. (8)-(11) has two merits. First, the metric transforms the observation y_i that are contaminated by the ISI into a new domain, composed by $z_{k,1}, z_{k,2}, z_{k,3}$ which are insensitive to ISI [17], [19], [20]. This suggests that the high-dimensional metric composed of such sub-metrics can inherently counteract the ISI, without the usages of the complex channel estimations for ISI compensation (as the coherent schemes do). Hence, this high-dimensional metric convert the detection with ISI into a binary detection framework, which enables potential designs of the non-coherent schemes.

Second, the usages of sets i.e., $\mathcal{N}_{\max}$, $\mathcal{N}_0$ and $\mathcal{N}_{\text{infx}}$ instead of the exact positions of the maximal value, the starting point and the inflexion point in Eqs. (9)-(11) help neutralize the sensitiveness to accurate synchronization. For instance, with a small synchronization error (that can be obtained via the technique in [29]), the exact position of the maximal value is still in the set $\mathcal{N}_{\max}$, making the sub-metric $z_{k,1}$ insensitive to the inaccurate synchronization. We will demonstrate the detection accuracy versus synchronization error in the Fig. 6 of Section V.

B. Distribution of High-dimensional Metric

Due to the central limit theorem (CLT), the sub-metrics $z_{k,1}$, $z_{k,2}$ and $z_{k,3}$ from Eqs. (9)-(11) that are the weighted summations of the observation y_i, can be regarded as Gaussian random variables (RVs), when the sample rate is large (e.g., $M = T_s/T_b \geq 50$). Hence, the d-dimensional $\mathbf{z}_k$ follows the multivariate normal distribution, i.e.,

$$\mathbf{z}_k \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), & \alpha_k = 1, \\ \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}), & \alpha_k = 0. \end{cases} \tag{12}$$

where $\boldsymbol{\mu}_j = \mathbb{E}(\mathbf{z}_k|\alpha_k = j)$ represents the mean of $\mathbf{z}_k$. $\boldsymbol{\Sigma}$ gives the covariance matrix.

Then, according to Eq. (12), the likelihood PDF of the d-dimensional metric $\mathbf{z}_k$ conditioned on the different information bits (i.e., $\alpha_k = 1$ or $\alpha_k = 0$) are:

$$\varphi(\mathbf{z}_k|\alpha_k = j) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}(\mathbf{z}_k - \boldsymbol{\mu}_j)^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}_k - \boldsymbol{\mu}_j)\right). \tag{13}$$

Here, it is noteworthy that the likelihood PDF in Eq. (13) also accounts for any designed d-dimensional metric $\mathbf{z}_k$ in Eq. (7), with its mean $\boldsymbol{\mu}_j$ and variance $\boldsymbol{\Sigma}$.

C. Decision Process

After specifying the form of the d-dimensional metric and its distributions, we then study the detection process. By adopting the Bayesian rule, the detection process is equivalent to compute and compare the posterior PDFs of $\mathbf{z}_k$ from different cases of $\alpha_k = 0$ and $\alpha_k = 1$, i.e.,

$$\frac{p(\alpha_k = 1|\mathbf{z}_k)}{p(\alpha_k = 0|\mathbf{z}_k)} \overset{\text{(a)}}{=} \frac{\varphi(\mathbf{z}_k|\alpha_k = 1)}{\varphi(\mathbf{z}_k|\alpha_k = 0)} \underset{\hat{\alpha}_k=0}{\overset{\hat{\alpha}_k=1}{\gtrless}} 1. \tag{14}$$

In Eq. (14), $p(\alpha_k|\mathbf{z}_k) = \varphi(\mathbf{z}_k|\alpha_k)\cdot\Pr\{\alpha_k\}$ represents the posterior PDF. (a) accounts for the assumption of the identically prior probability as $\Pr\{\alpha_k = 0\} = \Pr\{\alpha_k = 1\} = 0.5$.

Note that the detection process in Eq. (14) cannot be pursued, as one does not know the parameters (i.e., $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) of the likelihoods in Eq. (13). We will address this in Section IV.

D. Theoretical Performance Analysis

Premised on the construction of the high-dimensional metric $\mathbf{z}_k$, we then compute the theoretical BER, and then compare it with different non-coherent schemes.

1) Theoretical Decision Surface: The computation of the theoretical BER requires the decision surface from Eq. (14), which can be transformed as the logarithm difference i.e.,

$$\ln \varphi(\mathbf{z}_k|\alpha_k = 1) - \ln \varphi(\mathbf{z}_k|\alpha_k = 0) = 0, \tag{15}$$

Taken Eq. (13) into the left-hand of Eq. (15), the decision surface is computed as:

$$\mathcal{G}(\mathbf{z}_k) = 0, \tag{16}$$

with

$$\mathcal{G}(\mathbf{z}_k) = (\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_0^T)\boldsymbol{\Sigma}^{-1}\mathbf{z}_k - \frac{1}{2}(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0). \tag{17}$$

With the help of the decision surface in Eqs. (16)-(17), we then compute the theoretical BER.

2) Computation of BER: The theoretical BER of the d-dimensional $\mathbf{z}_k$ consists of two identical error probabilities, i.e.,

$$\begin{aligned}
P_{\mathrm{e}}(\mathbf{z}_k) =& \overbrace{\int\cdots\int}^{d}_{\mathcal{G}(\mathbf{z})\geqslant 0}\varphi(\mathbf{z}|\alpha_k = 0)d\mathbf{z} + \overbrace{\int\cdots\int}^{d}_{\mathcal{G}(\mathbf{z})<0}\varphi(\mathbf{z}|\alpha_k = 1)d\mathbf{z} \\
=& 2\overbrace{\int\cdots\int}^{d}_{\mathcal{G}(\mathbf{z}+\boldsymbol{\mu}_0)\geqslant 0}\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}\mathbf{z}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}\right)dz_1\cdots z_d \\
\overset{\text{(i)}}{=}& 2\overbrace{\int\cdots\int}^{d}_{\mathcal{G}(\boldsymbol{\Gamma}\cdot\mathbf{x}+\boldsymbol{\mu}_0)\geqslant 0}\frac{|\mathbf{J}|}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Lambda}^{-1}\mathbf{x}\right)dx_1\cdots x_d \\
\overset{\text{(ii)}}{=}& 2\overbrace{\int\cdots\int}^{d}_{\mathcal{G}(\boldsymbol{\Gamma}\cdot\boldsymbol{\Lambda}^{0.5}\cdot\mathbf{u}+\boldsymbol{\mu}_0)\geqslant 0}\frac{1}{\sqrt{(2\pi)^d}}\exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right)du_1\cdots u_d \\
\overset{\text{(iii)}}{=}& 2\cdot\Phi\left(-\frac{1}{2}\sqrt{(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)}\right).
\end{aligned} \tag{18}$$

The explanation of Eq. (18) is given as follows.

(i) As we observe that $\mathbf{z}^T\cdot\boldsymbol{\Sigma}\cdot\mathbf{z}$ holds for the quadratic form with $\boldsymbol{\Sigma}$ a d-dimensional symmetric matrix, we have $\mathbf{z} = \boldsymbol{\Gamma}\cdot\mathbf{x}$, and $\mathbf{z}^T\cdot\boldsymbol{\Sigma}^{-1}\cdot\mathbf{z} = \mathbf{x}^T\cdot\boldsymbol{\Lambda}^{-1}\cdot\mathbf{x}$, where $\boldsymbol{\Lambda} = diag\{\lambda_1, \cdots, \lambda_d\}$ is composed of d different eigenvalues of $\boldsymbol{\Sigma}$, and $\boldsymbol{\Gamma}$ is composed of the corresponding d normalized eigenvectors, such that $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\cdot\boldsymbol{\Lambda}\cdot\boldsymbol{\Gamma}^{-1}$. Hence, by replacing $\mathbf{z}$ with $\mathbf{x}$, the equation of (i) is achieved with the help of the Jacobian determinant, as

$$|\mathbf{J}| = \begin{vmatrix} \frac{\partial z_1}{\partial x_1} & \cdots & \frac{\partial z_1}{\partial x_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_d}{\partial x_1} & \cdots & \frac{\partial z_d}{\partial x_d} \end{vmatrix} = |\boldsymbol{\Gamma}| = 1. \tag{19}$$

(ii) We here use $\mathbf{u} = \boldsymbol{\Lambda}^{-0.5}\mathbf{x}$ to replace $\mathbf{x}$. Then, by noticing that $|\boldsymbol{\Sigma}| = |\boldsymbol{\Lambda}|$, we can obtain the equation of (ii).

(iii) Before we explain the establishment of (iii), we firstly prove the equality as follows:

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\cdot\Phi\left(az+b\right)dz \equiv \Phi\left(\frac{b}{\sqrt{1+a^2}}\right), \tag{20}$$

where

$$\Phi(x) \triangleq \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)dz. \tag{21}$$

This is because

$$\begin{aligned}
&\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\cdot\Phi\left(az+b\right)dz \\
=&\int_{-\infty}^{b}\frac{\partial\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\cdot\Phi\left(az+x\right)dz\right)}{\partial x}dx \\
=&\int_{-\infty}^{b}\frac{1}{\sqrt{2\pi(1+a^2)}}\exp\left(-\frac{x^2}{2(1+a^2)}\right)dx \\
=&\Phi\left(\frac{b}{\sqrt{1+a^2}}\right).
\end{aligned} \tag{22}$$

Then, by taking Eq. (20) back to (ii), we can prove the (iii) via

$$\begin{aligned}
&2\cdot\overbrace{\int\cdots\int}^{d}_{\mathcal{G}(\boldsymbol{\Gamma}\cdot\boldsymbol{\Lambda}^{0.5}\cdot\mathbf{u}+\boldsymbol{\mu}_0)\geqslant 0}\frac{1}{\sqrt{(2\pi)^d}}\exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{u}\right)du_1\cdots u_d \\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{d-1}^2}{2}\right)\cdot\int_{-\frac{\sum_{i=1}^{d-1}\xi_i\cdot u_i - B}{\xi_d}}^{+\infty} \\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{d-1}^2}{2}\right)\cdot\Phi\left(\frac{\sum_{i=1}^{d-1}\xi_i\cdot u}{\xi_d}\right. \\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{d-1}^2}{2}\right)\cdot\Phi\left(\frac{\xi_{d-1}}{\xi_d}u_{d-1}+\right. \\
=&2\cdot\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_1^2}{2}\right)\cdots\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u_{d-2}^2}{2}\right)\cdot\Phi\left(\frac{\sum_{i=1}^{d-2}\xi_i u_i}{\sqrt{\xi_d^2+\xi_d^2}}\right. \\
=&\cdots=2\cdot\Phi\left(-\frac{B}{\sum_{i=1}^{d}\xi_i^2}\right)=2\cdot\Phi\left(-\frac{0.5\boldsymbol{\xi}^T\cdot\boldsymbol{\xi}}{\boldsymbol{\xi}^T\boldsymbol{\xi}}\right)=2\cdot\Phi\left(-\frac{1}{2}\sqrt{\boldsymbol{\xi}^T\cdot\boldsymbol{\xi}}\right.
\end{aligned} \tag{23}$$

where $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_d]^T = \boldsymbol{\Lambda}^{-0.5}\boldsymbol{\Gamma}^{-1}\cdot(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$, and $B = 0.5\cdot\boldsymbol{\xi}^T\cdot\boldsymbol{\xi}$.

Hence, in Eq. (18), we give the general form of BER for any designed d-dimensional metric $\mathbf{z}_k$, as one can take the metric-related $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$ into Eq. (18). For instance, in our analysis of the MCvD, we can derive the theoretical BER of our $d = 3$ metric as well as the one of our previously proposed linear ($d = 1$) metric in [17], and their BER performances can be compared from the theoretical perspective.

3) Comparison among non-coherent schemes: We compare the BERs between the non-coherent schemes with metrics as a designed d-dimensional $\mathbf{z}_k$, and a $(d_1 \leq d)$-dimensional metric, i.e., $\mathbf{A}\cdot\mathbf{z}_k$. Here, we assume $rank(\mathbf{A}) = d_1$, otherwise its dimension can be reduced such that the determinant of its covariance matrix is positive, i.e., $|\mathbf{\Sigma}(\mathbf{A}\cdot\mathbf{z}_k)| > 0$. We give the comparison result as:

$$P_{\mathrm{e}}(\mathbf{z}_k) \leq P_{\mathrm{e}}(\mathbf{A}\cdot\mathbf{z}_k), \tag{24}$$

where the equality holds as $d_1 = d$.

The proof of the Eq. (24) is given as follows. Given Eq. (18), and the monotonically increasing property of $\Phi(x)$ with respect to x, the comparison between $P_{\mathrm{e}}(\mathbf{z}_k)$ and $P_{\mathrm{e}}(\mathbf{A}\cdot\mathbf{z}_k)$ can be converted as analyzing the quadratic values of $\boldsymbol{\mu}^T(\mathbf{\Sigma}^{-1} - \mathbf{A}^T(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T)^{-1}\mathbf{A})\boldsymbol{\mu}$ with $\boldsymbol{\mu} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ and $\mathbf{\Sigma}(\mathbf{A}\cdot\mathbf{z}_k) = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$. In this view, Eq. (24) is equivalent with $\mathbf{\Sigma}^{-1} - \mathbf{A}^T(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T)^{-1}\mathbf{A}$ being positive semidefinite.

(i) In the case of $rank(\mathbf{A}) = d_1 = d$, we have

$$\mathbf{\Sigma}^{-1} - \mathbf{A}^T(\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T)^{-1}\mathbf{A} = \mathbf{0} \tag{25}$$

which proves the equality of the Ineq. (24)

(ii) For $d_1 < d$, we consider only $d_1 = d-1$, and other cases $d_1 < d-1$ can be proved successively by replacing d as $d-1$. We firstly divide $\mathbf{A} = \mathbf{A}_1\cdot\mathbf{\Gamma}^{-1}$ (where $\mathbf{\Gamma}$ is the matrix of d eigenvectors of $\mathbf{\Sigma}$, and therefore $rank(\mathbf{\Gamma}) = d$), with $\mathbf{A}_1 = \mathbf{A}\cdot\mathbf{\Gamma}$. Then, the metric can be converted as $\mathbf{A}\cdot\mathbf{z}_k = \mathbf{A}_1\cdot\mathbf{\Gamma}^{-1}\mathbf{z}_k$ whereby the covariance matrix of $\mathbf{\Gamma}^{-1}\cdot\mathbf{z}_k$ is $\mathbf{\Sigma}(\mathbf{\Gamma}^{-1}\cdot\mathbf{z}_k) = \mathbf{\Lambda}$. Also, note that $\mathbf{A}_1 = \mathbf{\Pi}\cdot\mathbf{A}_2$, with $\mathbf{\Pi}$ the multiplications of elementary row transformations ($rank(\mathbf{\Pi}) = d-1$), and $\mathbf{A}_2 = [\mathbf{I}_{d_1\times d_1}\ \mathbf{a}]$ where $\mathbf{a} = [a_1, \cdots a_{d-1}]^T$. Hence, according to (i), with $P_{\mathrm{e}}(\mathbf{\Gamma}^{-1}\mathbf{z}_k) = P_{\mathrm{e}}(\mathbf{z}_k)$, and $P_{\mathrm{e}}(\mathbf{A}\cdot\mathbf{z}_k) = P_{\mathrm{e}}(\mathbf{\Pi}\cdot\mathbf{A}_2\cdot\mathbf{\Gamma}^{-1}\mathbf{z}_k) = P_{\mathrm{e}}(\mathbf{A}_2\cdot\mathbf{\Gamma}^{-1}\mathbf{z}_k)$, we only need to prove that $\mathbf{\Lambda}^{-1} - \mathbf{A}_2^T(\mathbf{A}_2\mathbf{\Lambda}\mathbf{A}_2^T)^{-1}\mathbf{A}_2$ is a positive semidefinite matrix. This can be illustrated as it has only non-negative eigenvalues:

$$\begin{aligned}
&\mathbf{\Lambda}^{-1} - \mathbf{A}_2^T\left(\mathbf{A}_2\mathbf{\Lambda}\mathbf{A}_2^T\right)^{-1}\mathbf{A}_2 \\
=&\mathbf{\Lambda}^{-1} - \mathbf{A}_2^T\left(diag(\lambda_1,\ldots,\lambda_{d-1})^{-1} - \zeta\cdot\boldsymbol{\varrho}\boldsymbol{\varrho}^T\right)\mathbf{A}_2 \\
=&\zeta\cdot\begin{bmatrix}\boldsymbol{\varrho}\boldsymbol{\varrho}^T & -\boldsymbol{\varrho}\\ -\boldsymbol{\varrho}^T & 1\end{bmatrix} \\
=&\zeta\cdot\mathbf{\Upsilon}\cdot diag\left(tr\left(\begin{bmatrix}\boldsymbol{\varrho}\boldsymbol{\varrho}^T & -\boldsymbol{\varrho}\\ -\boldsymbol{\varrho}^T & 1\end{bmatrix}\right), 0, \cdots 0\right)\cdot\mathbf{\Upsilon}^{-1} \\
=&\mathbf{\Upsilon}\cdot diag\left(\zeta\cdot tr\left(1+\boldsymbol{\varrho}^T\boldsymbol{\varrho}\right), 0, \cdots 0\right)\cdot\mathbf{\Upsilon}^{-1},
\end{aligned} \tag{26}$$

with $\boldsymbol{\varrho} = [a_1\sqrt{\lambda_d}/\lambda_1, \ldots, a_{d-1}\sqrt{\lambda_d}/\lambda_{d-1}]^T$, $\zeta = 1/(1+\lambda_d\sum_{i=1}^{d-1}a_i^2/\lambda_i)$, and $\mathbf{\Upsilon}$ the eigenvector matrix, which therefore proves the Ineq. (24).

The meaning of the Eq. (24) is explained as follows. First, it demonstrates that via using the same samples, the metric with higher dimension has the lower BER, i.e.,

$$\begin{aligned}
&P_{\mathrm{e}}(\mathbf{A}_{d\times kM}\cdot\mathbf{y}_{1:kM}) \\
&\quad< P_{\mathrm{e}}(\mathbf{A}_{(d_1<d)\times d}\cdot\mathbf{A}_{d\times kM}\cdot\mathbf{y}_{1:kM}).
\end{aligned} \tag{27}$$

This further indicates that our designed 3-dimensional metric $\mathbf{z}_k = [z_{k,1}, z_{k,2}, z_{k,3}]^T$ outperforms the previous linear one in [17] as $z_{k,1} + z_{k,2} + z_{k,3}$, i.e.,

$$P_{\mathrm{e}}(\mathbf{z}_k) < P_{\mathrm{e}}(z_{k,1} + z_{k,2} + z_{k,3}). \tag{28}$$

Secondly, the Eq. (24) indicates that the lower-bound of any designed metric $\mathbf{z}_k$ is the BER of the MAP under the known of the CIR, i.e.,

$$P_{\mathrm{e}}(\mathbf{z}_k) \geq P_{\mathrm{e}}(\mathbf{y}_{1:k}) \geq P_{\mathrm{e}}(\mathrm{MAP}), \tag{29}$$

as the Viterbi tracing algorithm can be pursued to reduce the BER by counteracting the effect of the ISI.

However, we should also notice two aspects. For one thing, the coherent MAP can only reach its theoretical BER if the CIR $h(t)$ is known. Otherwise, its performance depends directly on how accurate the estimation of the CIR is, which is difficult and will consume most of its computational resources [20]. Secondly, there is a resource-performance trade-off as we select and design the proper metric for the detection process. The performance of d-dimensional non-coherent scheme betters as the d grows larger, yet at the expense of both computational and storage complexities with respect to d, especially when we need to compute (or approximate) its d-dimensional multivariate likelihood PDFs in Eq. (13).

IV. Parzen-PNN realization

It is noteworthy that the detection process in Eq. (14) requires the computations of the likelihood PDFs conditioned on the mean $\boldsymbol{\mu}_j$ and the covariance matrix $\mathbf{\Sigma}$. Conventional channel estimation technique may fall into ether a long overhead or a relative poor acquisition of such parameters, especially for the covariance matrix $\mathbf{\Sigma}$. As such, it is demanding to resort to an alternative method that is capable of approximating the likelihood PDF in Eq. (13), but avoiding the complex estimation process on $\boldsymbol{\mu}_j$ and $\mathbf{\Sigma}$. And here come the Parzen window technique and its PNN based implementation.

A. Parzen Window technique

In essence, the Parzen-windowing technique [30] approximates the probability by defining a window (given the window size) and a function on this window (i.e. referred as the window function). Given a d-dimensional RV $\mathbf{z}$ with unknown distribution, Parzen-windowing approximates the corresponding PDF $p(\mathbf{z})$ by directly extracting the samples within the window function $\Pi(\mathbf{z})$, i.e. [21], [30],

$$p(\mathbf{z}) \simeq \frac{1}{N}\sum_{n=1}^{N}\frac{1}{\varsigma^d}\cdot\Pi\left(\frac{\mathbf{z}-\mathbf{z}^{(n)}}{\varsigma}\right), \tag{30}$$

where $\mathbf{z}^{(n)},\ n = 1, 2, ..., N$ is the samples, and ς denotes a smooth parameter that corresponds to the width of the window function. The accuracy of the approximation depends on the design of the window function $\Pi(\mathbf{z})$, which should also be a PDF in order to guarantee its approximated $p(\mathbf{z})$ as a PDF [22].

In the context of the MCvD applications, the aim is to approximate the likelihood functions $\varphi(\mathbf{z}_k|\alpha_k)$ via Eq. (30). Here, we adopt a Gaussian Parzen window of $\Pi(\mathbf{z})$, as we notice from Eq. (12) that the d-dimensional metric $\mathbf{z}_k$ follows a Gaussian distribution, i.e.,

$$\Pi(\mathbf{z}) = \frac{1}{(2\pi)^{d/2}}\exp\left(-\frac{1}{2}\mathbf{z}^T\mathbf{z}\right). \tag{31}$$

Then, by taking this window function $\Pi(\mathbf{z})$ into Eq. (30), we derive the approximated likelihood as:

$$\varphi(\mathbf{z}_k|\alpha_k) \backsimeq \frac{1}{N}\sum_{n=1}^{N}\frac{1}{(2\pi\varsigma^2)^{d/2}}\exp\left(-\frac{(\mathbf{z}_k-\mathbf{z}^{(n)})^T(\mathbf{z}_k-\mathbf{z}^{(n)})}{2\varsigma^2}\right). \tag{32}$$

This Parzen window technique has two obvious merits. First, it avoids the complex estimations of parameters, i.e., $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}$. By directly feeding the samples into the channel-independent Gaussian Parzen window function, the Parzen window technique approximates the likelihoods in Eq. (13) without the complex estimation of the channel parameters. Therefore, it has lower computational complexity, compared with the channel estimation algorithms that require the complex parameter acquisitions (especially the covariance matrix $\boldsymbol{\Sigma}$). Second, this approximated PDF approaches the theoretical perfectness when the number of samples grows infinity i.e., $N \to +\infty$, since:

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{(2\pi\varsigma^2)^{d/2}}\exp\left(-\frac{(\mathbf{z}_k-\mathbf{z}^{(n)})^T(\mathbf{z}_k-\mathbf{z}^{(n)})}{2\varsigma^2}\right)\\
&\overset{(a)}{=}\int_{\mathbb{R}^d} p(\mathbf{z}^*)\frac{1}{(2\pi\varsigma^2)^{d/2}}\exp\left(-\frac{(\mathbf{z}_k-\mathbf{z}^*)^T(\mathbf{z}_k-\mathbf{z}^*)}{2\varsigma^2}\right)d\mathbf{z}^*\\
&\overset{(b)}{=}\int_{\mathbb{R}^d}\varphi(\mathbf{z}^*|\alpha_k)\cdot p(\mathbf{z}_k|\mathbf{z}^*)d\mathbf{z}^*\\
&\overset{(c)}{=}\varphi(\mathbf{z}_k|\alpha_k).
\end{aligned} \tag{33}$$

In Eq. (33), (a) accounts for the mean expression of $1/(2\pi\varsigma^2)^{d/2}\exp(-(\mathbf{z}_k-\mathbf{z}^*)^T(\mathbf{z}_k-\mathbf{z}^*)/(2\varsigma^2))$ given the distribution of $\mathbf{z}^*$. (b) holds for the fact that the sampled $\mathbf{z}^* \sim \varphi(\mathbf{z}^*|\alpha_k)$, and the rest follows a Gaussian PDF conditioned on $\mathbf{z}^*$. The result of (c) resorts to the Chapman-Kolmogorov function [31], meaning that likelihood estimation is perfect.

B. Implementation of Parzen Window Technique

In order to implement the Parzen window technique to derive the approximated likelihoods in Eq. (32), and further pursue Eq. (14) for detection, we introduce the Parzen window based probabilistic neural networks [23]. The structure of the Parzen-PNN is illustrated in Fig. 3. From the input layer, the $(d=3)$-dimensional metric $\mathbf{z}_k$ is directly fed and serves as the Parzen-PNN input. The pattern layer is composed of the N previously computed metrics, i.e., $\mathbf{z}^{(n)} = \mathbf{z}_{k-n}$, labelled by the corresponding previous detection results $\alpha_n^* = \hat{\alpha}_{k-n}$. The output layer contains $C=2$ class representing 0-bit and 1-bit respectively.

1) Training process: The training process of the Parzen-PNN is provided in Algo. 1, which directly assigns the N values of pattern $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, ..., \mathbf{z}^{(N)}$ into pattern layer, and then labels each pattern as only one class of 0-bit and 1-bit via previous detection results. As such, the training process replaces the use of the predefined pilot sequence, with the previous data that can reflect the variation of the time-varying channel. We provide the detailed derivation of $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, ..., \mathbf{z}^{(N)}$ as well as the corresponding labels as follows.

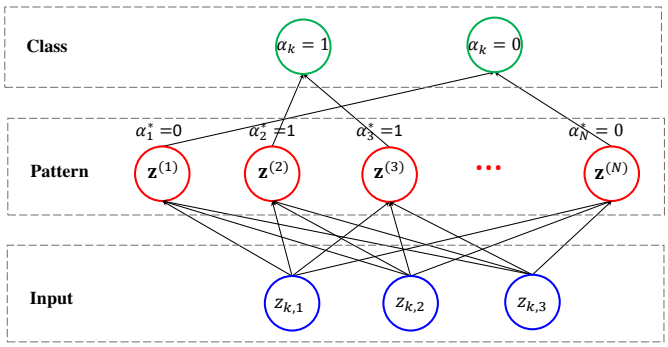

Fig. 3. Illustration of Parzen-PNN that consists of three layers, i.e., the input layer fed by the designed metric $\mathbf{z}_k$, the pattern layer trained by the labels $\alpha_1^*, \alpha_2^*, ..., \alpha_N^*$, and the class layer for classification.

Algorithm 1 Training Algorithm for kth symbol

Input: Previous data metric $\mathbf{z}_1 \cdots, \mathbf{z}_{k-1}$ with detection results $\hat{\alpha}_1, \cdots, \hat{\alpha}_{k-1}$.

1: **for** $n \in \{1, 2, ..., \min\{k-1, N\}\}$ **do**
2: Assign $\mathbf{z}^{(n)} = \mathbf{z}_{k-n}$, and $\alpha_n^* = \alpha_{k-n}$.
3: **end for**

Output: Return pattern values $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \cdots$, with labels $\alpha_1^*, \alpha_2^*, \cdots$.

(i) At the beginning when all signals are unlabelled, we use our previously proposed linear non-coherent scheme to label first N signals as 0-bit or 1-bit, i.e., [17]

$$\begin{aligned}
z_n^{\text{linear}} &= z_{n,1} + z_{n,2} + z_{n,3} \underset{\hat{\alpha}_n=0}{\overset{\hat{\alpha}_n=1}{\gtrless}} \eta_n, \quad 1 \le n \le N\\
\eta_n &= (1-a)\eta_{n-1} + a\frac{1}{n}\sum_{l=1}^{n} z_l^{\text{linear}}, \quad a \in [0.9, 0.99].
\end{aligned} \tag{34}$$

From Eq. (34), we construct $\mathbf{z}^{(n)} = [z_{n,1}, z_{n,2}, z_{n,3}]^T$ with label $\alpha_n^* = \hat{\alpha}_n$.

(ii) For any $k > N$, we use the previous detection results as the patterns, i.e., $\mathbf{z}^{(1)} = \mathbf{z}_{k-1}, \cdots, \mathbf{z}^{(N)} = \mathbf{z}_{k-N}$, with corresponding labels as $\alpha_1^* = \hat{\alpha}_{k-1}, \cdots, \alpha_N^* = \hat{\alpha}_{k-N}$.

In comparison with the conventional channel estimation algorithms that rely on the predefined pilot sequence to estimate the parameters, the training process of Parzen-PNN avoids the complex computation of $\boldsymbol{\mu}_j$, and $\boldsymbol{\Sigma}$, but directly extracts $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, ..., \mathbf{z}^{(N)}$ from previous data for further likelihood computations in the detection process. Also, the Parzen-PNN is relatively simple compared with the other neural network architectures, e.g., the back propagation (BP) networks, and the radial basis function networks (RBFN) [32].

2) Detection process: After above training process, the Parzen-PNN can be then adopted for signal detection. The detection algorithm is given by Algo. 2, which aims to compute the approximated likelihoods in Eq. (32) via the extracted $\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, ..., \mathbf{z}^{(N)}$. For each time-slot k, as we feed the d-dimensional metric $\mathbf{z}_k$ into the input layer, each pattern construct the Parzen window function by computing the inner

Algorithm 2 Detection Algorithm for kth symbol

Input: The trained Parzen-PNN, and the received signal at k time-step, i.e., $y_{(k-1)M+1}, \cdots, y_{kM}$.

1: Compute high dimensional metric $\mathbf{z}_k$ from Eqs. (9)-(11).
2: **for** $n \in \{1, 2, ..., N\}$ **do**
3: Compute $T(\mathbf{z}_k, \mathbf{z}^{(n)})$ via Eq. (35).
4: **end for**
5: Compute $\Delta_1(\mathbf{z}_k)$ and $\Delta_0(\mathbf{z}_k)$ via Eq. (36).
6: Derive $\hat{\alpha}_k$ by Eq. (37).

Output: Return $\hat{\alpha}_k$.

difference from its pattern value $\mathbf{z}^{(n)}$, i.e.,

$$\Pi\left(\frac{\mathbf{z}_k - \mathbf{z}^{(n)}}{\varsigma}\right) \propto T(\mathbf{z}_k, \mathbf{z}^{(n)}) = \exp\left(-\frac{(\mathbf{z}_k - \mathbf{z}^{(n)})^T(\mathbf{z}_k - \mathbf{z}^{(n)})}{2\varsigma^2}\right) \tag{35}$$

Then, the approximated likelihoods in Eq. (32) can be computed via the summation of each $T(\mathbf{z}_k, \mathbf{z}^{(n)})$ in the same category, i.e.,

$$\Delta_j(\mathbf{z}_k) = \sum_{\forall n,\ s.t.\ \alpha_n^* = j} T(\mathbf{z}_k, \mathbf{z}^{(n)}). \tag{36}$$

Finally, the detection can be pursued by the classification decision, as

$$\hat{\alpha}_k = \begin{cases} 1, & \Delta_1(\mathbf{z}_k) \geq \Delta_0(\mathbf{z}_k), \\ 0, & \Delta_1(\mathbf{z}_k) < \Delta_0(\mathbf{z}_k). \end{cases} \tag{37}$$

3) Smooth parameter: As is observed from Eq. (32), the smooth parameter ς^2 influences the accuracy of likelihood approximation and thereby the performance of signal detection. To be specific, a small value of ς^2 will produce a spiky approximation of the likelihood PDF in Eq. (13), giving rise to a rather curved decision surface and thereby causing extra miss detection. On the other hand, a large value of ς^2 will lead to an over-smoothing likelihood density, making it insensitive to the background noise. Here, we give two ideas to assign an appropriate ς^2.

Compared Eq. (32) with Eq. (13), it is intuitive to determine the smooth parameter as $\varsigma^2 = |\mathbf{\Sigma}|^{1/d}$. However, the estimation of the covariance matrix $\mathbf{\Sigma}$ requires great number of patterns as well as computational resources. In this view, we use the product of the trace of $\mathbf{\Sigma}$ as an approximation $|\mathbf{\Sigma}|$, and assign ς^2 as:

$$\varsigma^2 = \left(\prod_{i=1}^{d} \sigma_{z,i}^2\right)^{1/d}, \tag{38}$$

where $\sigma_{z,i}^2$ denotes the variance of ith sub-metric. In the proposed $d = 3$-dimensional noncoherent scheme, even if the distribution of the channel noise is unknown, we can still treat the $z_{k,1}$, $z_{k,2}$ and $z_{k,3}$ as Gaussian RVs, according to Eqs. (9)-(11). And their variances can be estimated via the labelled previous data.

C. Complexity Analysis

After a complete algorithm description, we analyze the computational complexity, by considering both the time and the storage consumption. In our discussions, we roughly measure this complexity by counting the total number of multiplication operations[2]. The complexity comparison among different schemes is provided in Table. I.

TABLE I
COMPARISON OF COMPLEXITY AMONG THREE DETECTION SCHEMES.

Methods/Complexity	Total complexity	Time consumption	Storage
Proposed HD non-coherent	$O(dN)$	$O(d)$	$O(dN)$
Linear non-coherent	$O(1)$	$O(1)$	$O(k)$
Coherent MAP	$O(k2^L + NL^2)$	$O(k2^L + NL^2)$	$O(k2^L)$

For the designed $(d = 3)$-dimensional non-coherent scheme at k time-step, the times of multiplications are $5 + dN = O(N + dN)$, which consists of 5 for construction of metric $\mathbf{z}_k$, and $d \times N$ for the detection process. Given the parallel structure of the Parzen-PNN, the exponential computation in Eq. (35) can be realized via the N parallel threads, which makes it have only d for time units. Thus, the total time consumption is $5 + d = O(d)$, at the expense of $O(dN)$ storage units. In this view, our scheme is suitable for real time communication, as we can spend just $O(d)$ time consumption for each informative bit detection.

For the previously proposed linear non-coherent scheme, at any k time-step, the computational complexity lies in (i) the computations of the three metrics in Eqs. (9)-(11), and (ii) the heuristic computation of the threshold η_k in Eq. (6). As such, the times of multiplications can be computed as $8 = O(1)$. Given that the computation of η_k requires the summations of $c_1^{\text{linear}}, \cdots, c_k^{\text{linear}}$, the storage needed for the linear non-coherent scheme is measured as $O(k)$.

Then, we compute the computational complexity of the coherent MAP scheme. The coherent MAP consists of the channel estimation, and the signal detection. For the Viterbi tracing detection process in Eq. (5), finding the maximum posterior density for current k time-step requires the recursive computations from $\max_{u_{k-1}} p(u_k|u_{k-1})p(\tilde{y}_k|u_k)$ to $\max_{u_1} p(u_2|u_1)p(\tilde{y}_2|u_2)$, with $u_k = \sum_{l=k-L+1}^{k} Q \cdot \tilde{h}_{l-k+1} \cdot \alpha_l$ for $k \geq L$, and $u_k = \sum_{l=1}^{k} Q \cdot \tilde{h}_{l-k+1} \cdot \alpha_l$ for $1 \leq k < L$ Hence, for each $L \leq k' \leq k$, 2^L likelihood densities will be computed for enumerating all potential $\boldsymbol{\alpha}_{k'-L+1:k'} \in \{0, 1\}^L$. For $1 \leq k' < L$, $2^{k'}$ likelihoods are computed. Given from Eq. (3), each logarithm-form likelihood PDF includes one logarithm computation. As such, the total number of multiplications for the MAP detection can be measured as $(k - L + 2) \times 2^L - 2 = O(k2^L)$. For the channel estimation algorithm, we adopt the ML channel estimation approach in [10], whereby a predefined pilot sequence of length N is used to estimate the CSI. To do so, three matrix (of size $L \times (N - L + 1)$) multiplications, and a multiplication of such matrix with a $(N - L + 1) \times 1$ vector are used, which requires $(N - L + 1) \times (2L^2 + L)$ times of multiplications. Combined with the detection process, the total number of multiplications for coherent MAP with ML channel estimation is $(k - L + 2) \times 2^L - 2 + (N - L + 1) \times (2L^2 + L) =$

[2]Here, the exponential (or logarithm) operation is treated as several multiplications according to Taylor expansion.

$O(k \times 2^L + N \times L^2)$. For the storage needed, the Viterbi tracing requires k observations, L CSI, as well as $2^L(2^{k'})$ likelihood densities for each $L \leq k' \leq k(1 \leq k' < L)$. As such, the storage required for the coherent MAP is $(k - L + 2) \times 2^L - 2 + L + k = O(k \times 2^L)$.

D. Implementation via Biological Circuits

Finally, we provide an outline of the biological design of the nano-receiver. The outline includes the feasibility analysis on both the sensory units for sensing the signaling molecules, and the proposed high-dimensional non-coherent scheme for signal detection.

For the sensory units, one can resort to the existing alcohol sensor in [33], and magnetic coils of the susceptometer in [34] respectively, based on which two MCvD systems using corresponding molecules have been designed. For the Cell-to-cell communications we can rely on the Ca^{2+} signaling, which is central to a multitude of biological processes including neuronal signaling. The concentration level of Ca^{2+} ions can be monitored using fluorescent Ca^{2+} indicators. For example, lightâĂŞexcitable Calcium Green-1 has been used to image neuronal activity [35]. The binding of Ca^{2+} ion to an indicator can lead to an increase fluorescence (quantum yield) or a wavelength shift that can be monitored optically, i.e., [35]

$$Ca^{2+} + B \leftrightarrows CaB$$

where B is the binding site, and CaB accounts for the calcium-bound complex. As such, by setting the calcium ion as the messenger molecules to convey the OOK information, the man-made nano-receiver fluorescent indicator can report on the number of molecules.

We then consider the biological design of the proposed detection scheme. The study in [36] analyzed the protein-based signaling within biological cells. It is demonstrated that the fundamental motif in all biological signaling is underpinned by the protein phosphorylation and dephosphorylation (referred as the cascade cycles). In the cascade cycles, the input of the protein phosphorylation transits the original state (0-state) into a new state (1-state), while the input of the protein dephosphorylation reverses such process. As such, the cascade cycles enable to construct various control and computational digital circuits (e.g., the RS-register, the AND, and NAND gates, the adder and the multiplier [36]). Given that the proposed high-dimensional non-coherent scheme includes only the adder and the multiplier (the exponential operation can be approximated by numbers of multiplier and adder according to Taylor expansion), the study in [36] thereby provides a great promise for implementation via the biological designed circuits.

V. Numerical Simulations

In the following analysis, the performance of our proposed high-dimensional non-coherent detection scheme (abbreviated as HD non-coherent) will be evaluated, in terms of the BER. First, we examine the performance of the Parzen-PNN realization by considering the smoothing parameter ς^2, and the number of the patterns N respectively. Then, we analyze

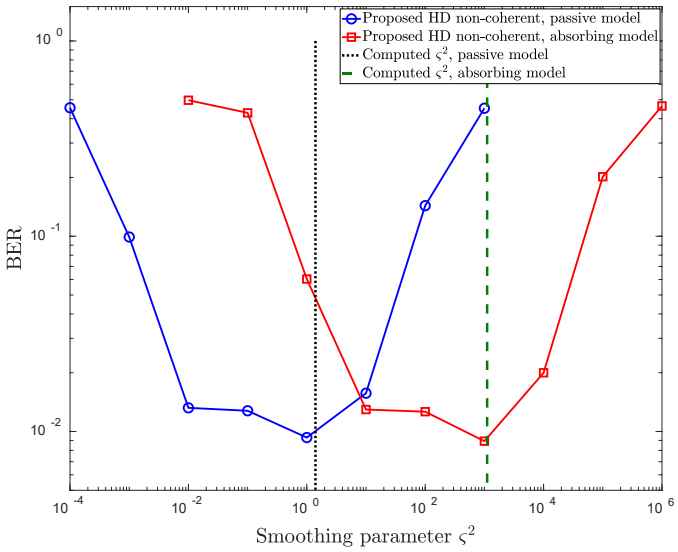

Fig. 4. Illustration of the importance of the smooth parameter on Parzen-PNN realization of the proposed HD non-coherent scheme.

the BER versus the number of samples per symbol, i.e., M. After that, the detection accuracy to synchronization error is tested. Finally, the comparisons between our proposed HD non-coherent scheme, the linear non-coherent scheme, and the state-of-the-art MAP are pursued, with respect to various SNR and the bit interval T_b that represents the intensity of the ISI.

The involved parameters in this simulation are configured according to [37], which are typical for nano-scale molecular communication between small biological cells [38]. As far as both the passive and the absorbing properties of the nano-receiver are concerned, we consider a spherical shape of the receiver with its radius as $R = 0.225\mu$m, and thus its volume as $V = 4/3\pi R^3 = 4.771 \times 10^{-2}(\mu\text{m})^3$. The sample rate $T_s = M \cdot T_b$ varies from the bit interval T_b, but we assure $M = 50$ samples for each bit detection. For the diffusive channel, we assign the diffusive coefficient as $D = 5 \times 10^3(\mu\text{m})^2/\text{s}$, the communication distance between the transmitter and the receiver as $r = 2\mu$m, the drift velocity as $v = 10^{-3}$m/s, and the degradation factor for enzyme effect as $\beta = 100$/s.

A. Performance of Parzen-PNN

The performance of the Parzen-PNN realization is illustrated in Fig. 4 and Fig. 5, as we fix the SNR $=$ 5dB, and the $T_b = 3 \times 10^{-4}$s.

We can firstly observe from the Fig. 4 that the smoothing parameter ς^2 plays an important role on the BER of the Parzen-PNN. Take the passive mode as an example. As ς^2 increases from 10^{-4} to 10^4, the BER decreases from 0.4, till reaching its optimal value (i.e., nearly 10^{-2}), and then begins to rebound. This is because that ς^2 reflects the intensity of variance $\mathbf{\Sigma}$ when approximating the likelihood PDF in Eq. (13) by using Eq. (32), and therefore neither a larger nor a smaller value of ς^2 performs well. More importantly, it is noticed in Fig. 4 that, the computation of $\varsigma^2 = (\sigma_{z,1}^2 \cdot \sigma_{z,2}^2 \cdot \sigma_{z,3}^2)^{1/3} = 4.65$ via Eq. (38) ensures the closeness to the optimal BER. This suggests

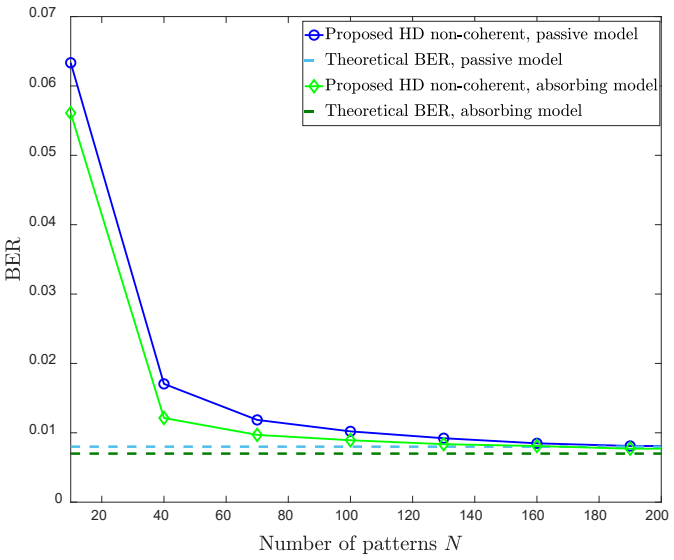

Fig. 5. Illustration of the BER convergence as the number of patterns N grows.

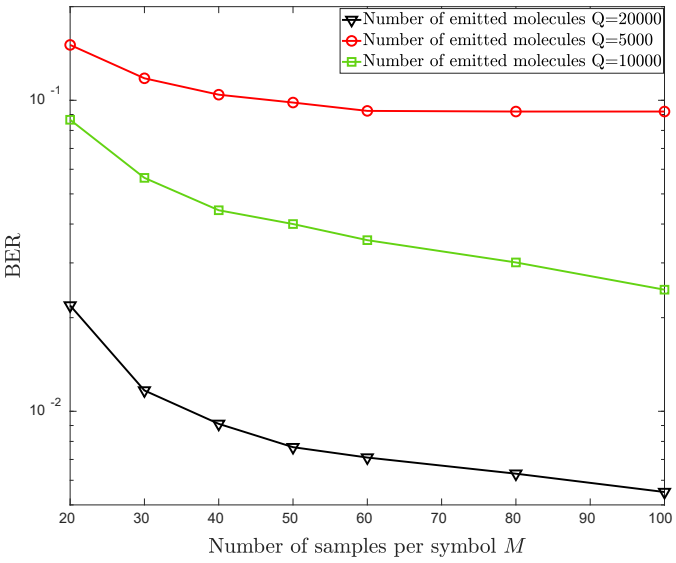

Fig. 6. Illustration of BER versus the number of samples per symbol M for the proposed HD non-coherent scheme.

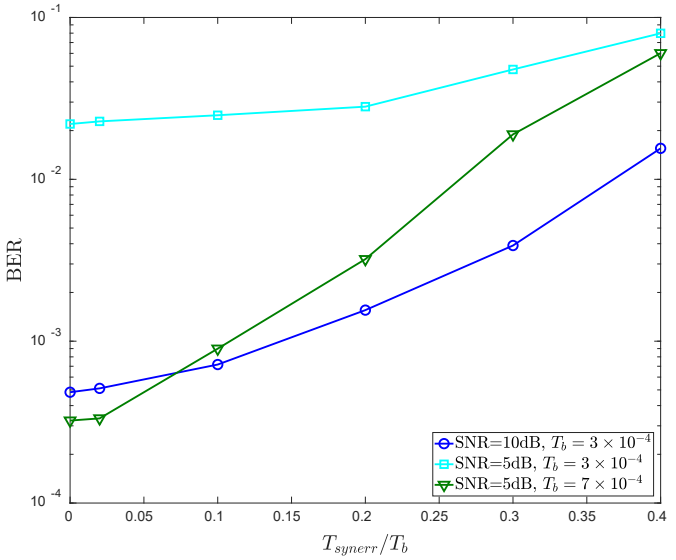

Fig. 7. Illustration of the influence of synchronization error on the proposed HD non-coherent scheme. A tolerant BER can be reached when the synchronization has a small error (e.g., $< 10\%$).

that a good performance of the Parzen-PNN realization can be achieved when choosing and computing an appropriate smoothing parameter ς^2.

Then, we demonstrate that given a proper smoothing parameter, the theoretical performance provided in Eq. (18) can be approached, as we increase the number of the patterns, i.e., N. It is observed from the Fig. 5 that the number of the patterns determines the BER performance of the Parzen-PNN realization. For example, with the increase of the number from 10 to 200, the BERs of both the absorbing and passive models converge to their theoretical limits (i.e., 0.06 for the absorbing and 0.07 for the passive). This verifies the correctness of Eq. (33) that proves the perfect approximations of the Parzen approximation as N grows to infinity.

With the help of Fig. 4 and Fig. 5, we demonstrate that the Parzen-PNN realization can approach the theoretical BER performance in the absence of any knowledge of the specific channel models. This indicates that our proposed HD non-coherent detection scheme is suitable for signal detection in MCvD. We will illustrate this in the following parts.

B. BER versus Number of Samples

The BER versus the number of samples for each symbol (i.e., M) is provided in Fig. 6, in which we examine three cases with different number of emitted molecules (i.e., $Q = 5000$, $Q = 10000$, and $Q = 20000$) per symbol. Here, the number of emitted molecules Q indicates two facts. First, given the Poisson distribution of the observation y_i in Eq. (3), a larger Q gives rise to a larger SNR. This can be demonstrated in Fig. 6, where the proposed HD non-coherent scheme with a larger Q (e.g., $Q = 20000$) has a lower BER than those with smaller Q (e.g., $Q = 10000$ and $Q = 5000$). Furthermore, as we increase the sample size (see $Q = 5000$ for example), the BER saturates to a lower limit. This is because when the number of emitted molecules Q is small, the less variety of molecules' movement leads to oversampling of the same physical state. As such, we also demonstrate via Fig. 6 that our selection of $M = 50$ is suitable for the proposed HD non-coherent scheme, as $M = 50$ provides lower BER but is still not in the region of BER convergence.

C. Influence of Synchronization error

The sensitiveness of the proposed HD non-coherent scheme to synchronization error is illustrated in Fig. 5. The synchronization error is measured via the percentage of the deviation to symbol interval, i.e., T_{synerr}/T_b. We test three cases with different SNR and T_b (that representing the intensities of the ISI). It is seen that the BER grows as the synchronization error increases. For instance, when the error reaches 40%, the

BER of the case SNR=5dB and $T_b = 7 \times 10^{-4}$ increases from 2×10^{-4} to 5×10^{-2}. This result suggests that the performance of the proposed HD non-coherent scheme is affected by the accuracy of the synchronization error.

More importantly, it is noteworthy that the proposed HD non-coherent scheme can also provide a tolerant BER if the synchronization error is small. For example, within a small area of synchronized error, i.e., $< 10\%$, the BER from the case SNR=10dB $T_b = 3 \times 10^{-4}$ increases from 5×10^{-4} to 7×10^{-4}, which indicates the low sensitiveness to the synchronization error. The reason for this low sensitiveness lies in the construction of the metrics in Eqs. (9)-(11), which use the average of the neighbouring values to extract the transient features. As such, even if the synchronization has a small error, the extractions of rising edge, the inflexion, and the signal difference between two adjacent symbols are still reliable for further detection process.

D. Performance Comparisons

The BER comparisons between the proposed HD noncoherent scheme, the linear noncoherent scheme, and the state-of-the-art MAP [11] equipped with the ML channel estimation [10] are illustrated in Fig. 7 and Fig. 8, where Fig. 7 gives the comparison with various SNR, and Fig. 8 shows the results affected by the intensities of ISI. The coherent MAP with the perfect knowledge of the CSI is provided as benchmark. Here, the SNR for this simulation is defined as the ratio of the power of the signal to the variance of the noise. Given the Poisson distribution of the observation y_i in Eq. (3), the power of the signal can be measured via the average of $s_i = \sum_{l=k-L+1}^{k} Q \cdot h_{i-(l-1)M} \cdot \alpha_l$, i.e., $\mathbb{E}(s_i)$, and the variance of noise for Poisson distribution is $\mathbb{E}(s_i) + \bar{\epsilon}_{\text{int}}$. As such, the SNR is expressed as: [26]

$$\text{SNR} = \frac{\mathbb{E}(s_i)^2}{\mathbb{E}(s_i) + \bar{\epsilon}_{\text{int}}}. \tag{39}$$

1) BER versus SNR: The BER performance with respect to the SNR is shown in Fig. 8, in which Fig. 8-(a) considers the case with signal-independent noise, and Fig. 8-(b) presents the case with signal-dependent noise. It is intuitive that the BERs of all the schemes are affected by the SNR. For instance, the BER of the MAP is deteriorated from 8×10^{-3} to 0.47, as the SNR decreases from 20dB to -5dB. Also, our proposed two non-coherent schemes perform badly in the region of the low intensities of the SNR.

Then, it is seen that our proposed HD non-coherent scheme is also suitable for the case with the signal-dependent noise. In Fig. 8-(b), the BER of the proposed scheme with signal-dependent noise decreases from the 0.2 to 5×10^{-4} when the SNR increases from -5dB to 10dB, which is similar with the performance provided in the case of signal-independent noise in Fig. 8-(a).

More importantly, we can observe that our proposed HD non-coherent scheme outperforms the state-of-the-art MAP and the linear noncoherent method. For instance, the BER of our HD non-coherent scheme is 6×10^{-4}, as the SNR=10dB, which is smaller than that of the MAP scheme (i.e., 0.1) and the linear scheme (i.e., 8×10^{-2}). The reason can be summarized as two aspects. **1)** Our proposed scheme relies on the high-dimensional construction of the three feature metrics, rather than their linear combination, capable of providing a larger SNR, and therefore can obtain a better performance given the Eq. (24). **2)** Under the assumption of the unknown channel models, the performance of the state-of-the-art MAP scheme is restricted by the accuracy of CSI estimation for ISI compensation, which, if has errors, will subsequently deteriorate the computation of the likelihood PDF, therefore leading to erroneous results for signal detection. By contrast, our proposed HD non-coherent scheme uses metric that can inherently counteract the ISI via the exploration of the transient features, and thereby capable of reaching a remarkable performance in the blind detection scenarios.

2) BER versus ISI: The performance of the BER corresponding to the ISI is provided in Fig. 9, in which Fig. 9-(a) considers the case with signal-independent noise, and Fig. 9-(b) shows the case with signal-dependent noise. It is shown that the intensities of the ISI influence the BERs of these schemes. For example, with the decrease of the symbol interval T_b from 1.1×10^{-3}s to 3×10^{-4}, the ISI becomes stronger, and therefore deteriorates the BER of the proposed HD non-coherent scheme from 9×10^{-4} to 8×10^{-2}.

Then, we can observe from Fig. 9-(b) that the detection performance of the proposed HD non-coherent scheme is also well-behaved for the case with signal-dependent noise. Similar to the BER in the case of signal-independet noise in Fig. 9-(a), the BER of the propose scheme reduces from 8×10^{-2} to 9×10^{-4}, as the symbol interval T_b grows from 3×10^{-4} to 11×10^{-4}.

When we make a comparison among the three schemes, we can see that the proposed HD non-coherent scheme performs better than the state-of-the-art MAP and the linear non-coherent scheme. For instance, the BER of our proposed HD non-coherent scheme varies from 8×10^{-2} to 9×10^{-4} as the T_b grows from 3×10^{-4} to 1.1×10^{-3}, smaller than the values of the linear scheme that ranges from 1.1×10^{-1} to 10^{-2}, not to mention the poor performances derived from the MAP method. This great advantage of our proposed scheme is attributed to following perspectives. As we have proved in the Eq. (24), the HD non-coherent has a lower BER as opposed to the scheme with lower dimensional metric. Besides, the advantage of blind detection of our scheme makes it possible to approximate the likelihood density and thereby leads to a more reliable detection results, as opposed to the MAP in [11] that will be harmed under the unknown CSI.

VI. Conclusion

In most sequential molecular signal detection scenarios, information recovery without explicit channel models is challenging. Current research challenges lie in how to deal with the background noise the ISI caused by the long-tail nature of the channel response and its dynamics. In this paper, we extend and improve our linear metric combining non-coherent detection method by designing a high-dimensional metric and proposing the Parzen-PNN implementation. We first construct

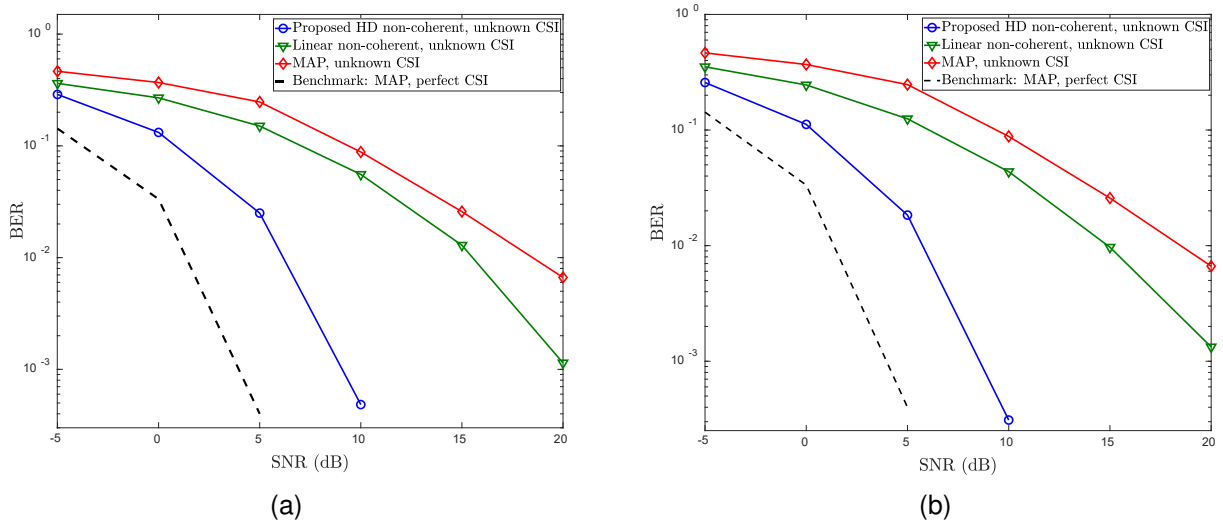

Fig. 8. BER comparisons among different schemes with respect to the changes of the SNR. (a) provides the scenario with signal independent noise. (b) presents the case with signal dependent noise. It is seen that in both cases, the proposed HD non-coherent scheme outperforms the MAP with unknown CSI and linear non-coherent method.

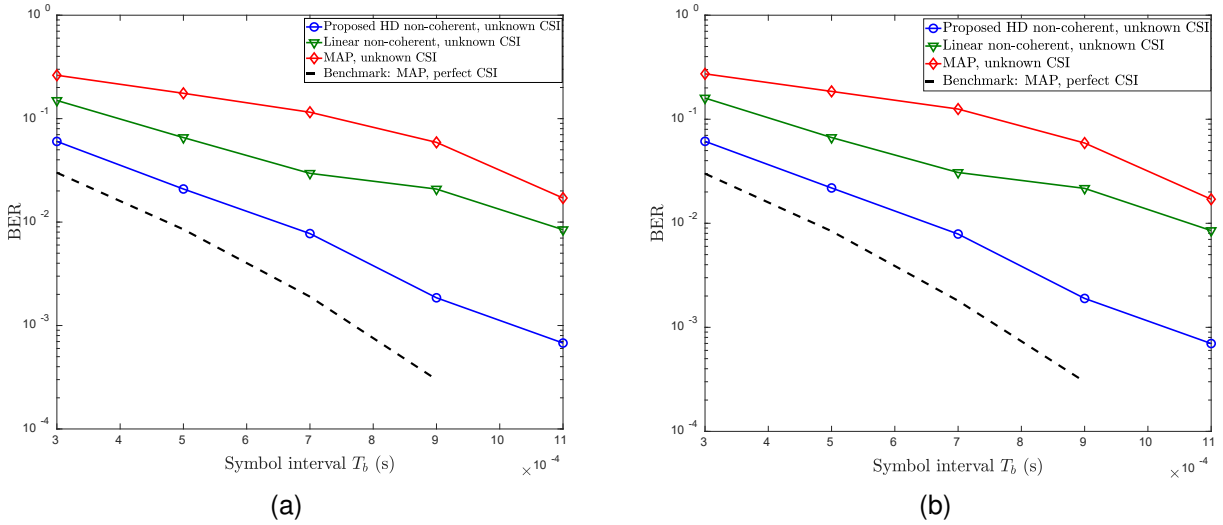

Fig. 9. BER comparisons among different schemes versus the changes of the symbol interval T_b that reflect the intensities ISI. (a) illustrates the scenario with signal independent noise. (b) gives the case with signal dependent noise. It is observed that in both cases, the proposed HD non-coherent scheme outperforms the MAP with unknown CSI and linear non-coherent method.

the high-dimensional metric by re-designing the sub-metrics to be robust against various ISI intensities. Then, we compute the theoretical BER bound of the high-dimensional metric and prove to have a lower BER as opposed to the linear scheme. By considering the blind detection scenarios, we adopt the Parzen-PNN implementation to estimate the likelihood PDF and then detect the information bits. In comparison with the state-of-the-art MAP and previous linear schemes, our newly proposed high-dimensional non-coherent scheme gains an average of 10dB in SNR given the fixed BER as 10^{-4}. This generalizable technique provides promising pathways for future research in adverse molecular or biological channels.

REFERENCES

[1] L. J. Kahl and D. Endy, "A survey of enabling technologies in synthetic biology," *Journal of Biological Engineering*, vol. 7, no. 1, p. 13, 2013.

[2] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The internet of bio-nano things," *IEEE Communications Magazine*, vol. 53, no. 3, pp. 32–40, 2015.

[3] P. Wu, C. Cheng, C. D. Kaddi, J. Venugopalan, R. Hoffman, and M. D. Wang, "ÂŰomic and electronic health record big data analytics for precision medicine," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 2, pp. 263–273, 2017.

[4] W. Guo, C. Mias, N. Farsad, and J. Wu, "Molecular versus electromagnetic wave propagation loss in macro-scale environments," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 1, no. 1, pp. 18–25, 2015.

[5] M. M. Al-ZuÂŠbi, A. S. Mohan, and S. S. H. Ling, "Impact of reactive obstacle on molecular communication between nanomachines," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 4468–4471.

[6] H. B. Yilmaz, G. Suk, and C. Chae, "Chemical propagation pattern for molecular communications," *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 226–229, 2017.

[7] H. B. Yilmaz, A. C. Heren, T. Tugcu, and C. Chae, "Three-dimensional channel characteristics for molecular communications with an absorbing

receiver," *IEEE Communications Letters*, vol. 18, no. 6, pp. 929–932, 2014.

[8] A. O. Bicen and I. F. Akyildiz, "Interference modeling and capacity analysis for microfluidic molecular communication channels," *IEEE Transactions on Nanotechnology*, vol. 14, no. 3, pp. 570–579, 2015.

[9] S. Kadloor, R. S. Adve, and A. W. Eckford, "Molecular communication using brownian motion with drift," *IEEE Transactions on NanoBioscience*, vol. 11, no. 2, pp. 89–99, 2012.

[10] V. Jamali, A. Ahmadzadeh, C. Jardin, H. Sticht, and R. Schober, "Channel estimation for diffusive molecular communications," *IEEE Transactions on Communications*, vol. 64, no. 10, pp. 4238–4252, 2016.

[11] D. Kilinc and O. B. Akan, "Receiver design for molecular communication," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 12, pp. 705–714, 2013.

[12] A. Noel, K. C. Cheung, and R. Schober, "Optimal receiver design for diffusive molecular communication with flow and additive noise," *IEEE Transactions on NanoBioscience*, vol. 13, no. 3, pp. 350–362, 2014.

[13] W. Guo, T. Asyhari, N. Farsad, H. B. Yilmaz, B. Li, A. Eckford, and C. Chae, "Molecular communications: channel model and physical layer techniques," *IEEE Wireless Communications*, vol. 23, no. 4, pp. 120–127, 2016.

[14] V. Jamali, N. Farsad, R. Schober, and A. Goldsmith, "Non-coherent detection for diffusive molecular communication systems," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2515–2531, 2018.

[15] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *IEEE Transactions on NanoBioscience*, vol. 13, no. 1, pp. 31–43, 2014.

[16] Y. J. Cho, H. B. Yilmaz, W. Guo, and C. Chae, "Effective enzyme deployment for degradation of interference molecules in molecular communication," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.

[17] B. Li, C. Zhao, and W. Guo, "Non-linear signal detection for molecular communications," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6.

[18] V. Jamali, A. Ahmadzadeh, N. Farsad, and R. Schober, "Constant-composition codes for maximum likelihood detection without csi in diffusive molecular communications," *IEEE Transactions on Communications*, vol. 66, no. 5, pp. 1981–1995, 2018.

[19] B. Li, M. Sun, S. Wang*, W. Guo, and C. Zhao, "Low-complexity noncoherent signal detection for nanoscale molecular communications," *IEEE Transactions on NanoBioscience*, vol. 15, no. 1, pp. 3–10, 2016.

[20] B. Li, M. Sun, S. Wang, W. Guo, and C. Zhao, "Local convexity inspired low-complexity noncoherent signal detector for nanoscale molecular communications," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2079–2091, 2016.

[21] B. Li, Z. Zhou, D. Li, and W. Zou, "A novel parzen probabilistic neural network based noncoherent detection algorithm for distributed ultra-wideband sensors," *Journal of Network and Computer Applications*, vol. 34, no. 6, pp. 1894–1902, 2011.

[22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[23] D. F. Specht, "Probabilistic neural networks," *Neural networks*, vol. 3, no. 1, pp. 109–118, 1990.

[24] J. Purvis and G. Lahav, "Encoding and decoding cellular information through signaling dynamics," *Cell*, vol. 152, no. 5, pp. 945 – 956, 2013.

[25] A. Noel, Y. Deng, D. Makrakis, and A. Hafid, "Active versus passive: Receiver model transforms for diffusive molecular communication," in *2016 IEEE Global Communications Conference (GLOBECOM)*, 2016, pp. 1–6.

[26] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication-a tutorial review," *arXiv preprint arXiv:1812.05492*, 2018.

[27] T. Deneux, A. Kaszas, G. Szalay, G. Katona, T. Lakner, A. Grinvald, B. Rózsa, and I. Vanzetta, "Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo," *Nature communications*, vol. 7, p. 12190, 2016.

[28] Z. Wei, W. Hu, M. Zhang, D. Han, B. Li, and C. Zhao, "Viterbi estimation on the finite-state markov ultra-violet channels," in *Asia Communications and Photonics Conference*. Optical Society of America, 2017, pp. Su2A–19.

[29] H. ShahMohammadian, G. G. Messier, and S. Magierowski, "Blind synchronization in diffusion-based molecular communication channels," *IEEE Communications Letters*, vol. 17, no. 11, pp. 2156–2159, 2013.

[30] E. Parzen, "On estimation of a probability density function and mode," *The annals of mathematical statistics*, vol. 33, no. 3, pp. 1065–1076, 1962.

[31] A. H. Jazwinski, *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[32] S. S. Haykin, *Adaptive filter theory*. Pearson Education India, 2008.

[33] N. Farsad, W. Guo, and A. W. Eckford, "Tabletop molecular communication: Text messages through chemical signals," *PloS one*, vol. 8, no. 12, p. e82935, 2013.

[34] H. Unterweger, J. Kirchner, W. Wicket, A. Ahmadzadeh, D. Ahmed, V. Jamali, C. Alexiou, G. Fischer, and R. Schober, "Experimental molecular communication testbed based on magnetic nanoparticles in duct flow," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[35] R. Brette and A. Destexhe, *Handbook of neural activity measurement*. Cambridge University Press, 2012.

[36] H. M. Sauro and B. N. Kholodenko, "Quantitative analysis of signaling networks," *Progress in biophysics and molecular biology*, vol. 86, no. 1, pp. 5–43, 2004.

[37] Y. Fang, A. Noel, N. Yang, A. W. Eckford, and R. A. Kennedy, "Convex optimization of distributed cooperative detection in multi-receiver molecular communication," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 3, no. 3, pp. 166–182, 2017.

[38] B. Alberts, D. Bray, K. Hopkin, A. D. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential cell biology*. Garland Science, 2013.

2019-12-13

High-dimensional metric combining for non-coherent molecular signal detection

Wei, Zhuangkun

IEEE

Wei Z, Guo W, Li B, et al., (2020) High-dimensional metric combining for non-coherent molecular signal detection. IEEE Transactions on Communications, Volume 68, Issue 3, March 2020, pp. 1479-1493

https://doi.org/10.1109/TCOMM.2019.2959354